J. HYM. RES.
1(1), 1992 pp. 63-79

The Application of Nucleotide Sequence Data to Phylogeny of the Hymenoptera: A Review

S.A. Cameron, J.N. Derr, A.D. Austin, J.B. Woolley and R.A. Wharton

(SAC) Department of Biology, Washington University, St. Louis, Missouri 63130-4899 U.S.A.[1]; (JND, JBW, RAW) Department of Entomology, Texas A & M University, College Station, Texas 77843-2475 U.S.A.; (ADA) Department of Crop Protection, Waite Campus, University of Adelaide, Glen Osmond, S.A. 5064 Australia.

Abstract.—The application of molecular sequence data to studies on the phylogeny of the Hymenoptera are reviewed, with special attention given to the relationships among the higher levels of the Order. Methods for obtaining sequence information from nuclear-encoded ribosomal RNA (rRNA) and mitochondrial rRNA and protein-coding genes are described. Techniques for alignment and phylogenetic analysis of sequences are discussed, as are issues associated with the selection of outgroups. Recent molecular investigations of hymenopteran phylogeny at several taxonomic levels are discussed to illustrate the application of methods and analytical procedures.

The use of DNA sequence data for systematics is recent and controversial. The controversies are not about whether nucleotide sequences are appropriate for reconstructing phylogenetic history but rather, how they should be used. Therefore, the springboard for our review is not a justification of the relative merits of sequence data over the application of other techniques for phylogenetic analysis (for this see Hillis and Moritz 1990), instead we begin with a discussion of the areas of controversy that have arisen with the use of DNA sequences for phylogenetic analysis. We review each of these issues and make recommendations based in part on our own experiences with collecting and analyzing DNA sequences of Hymenoptera.

Differences of opinion have arisen over aspects of sequence data collection and analysis, including (1) the appropriate genes (or gene fragments) to be sequenced and their use for different levels of inference; (2) methods of data acquisition; (3) methods of alignment, character weighting, and tree-building; (4) assumptions (or the lack thereof) of the models of nucleotide evolution; (5) consideration of molecular secondary structure and the degree to which it can bias interpretation of sequence data for phylogenetic reconstruction; and (6) appropriate statistical analyses for estimating the reliability of molecular phylogenies. Each of these issues confronts all systematists who wish to approach phylogenetic reconstruction from a molecular perspective, altogether a non-trivial pursuit for beginner and experienced alike.

This paper arose from the symposium 'Phylogeny of the Hymenoptera', which was featured during the 2nd Quadrennial meeting of the International Society of Hymenopterists, held in August, 1991 in Sheffield, England. Three contributions in the symposium presented results of phylogenetic analyses using DNA sequences. It become clear at this meeting that many of our audience were unfamiliar with the use of sequence data for systematic studies. In the future, systematists will have to interpret critically the results from molecular studies in order to compare them effectively with their own investigations based on morphology or other types of data. Therefore, we thought it worthwhile to review the subject of molecular phylogeny with particular reference to the Hymenoptera. To remain faithful to the theme of the symposium, we primarily restrict our discussion in this review to questions of higher level phylogeny, that is, to the tribal level or above. However, we include a single study of relationships at the species level. Given that little has been published on comparative DNA sequences for phylogenetic reconstruction of the Hymenoptera (but see Cameron 1991; Garnery et al. 1991; Sheppard and McPheron 1991), we rely heavily on our own investigations of sequence comparisons of the small (18S) subunit ribosomal RNA gene (rRNA) and the large (16S) rRNA gene encoded by the mitochondrial genome (mtDNA). For a general review of the field of molecular systematics we recommend two

[1]Current address: Department of Biological Sciences, University of Arkansas, Fayetteville, Arkansas 72701, U.S.A.

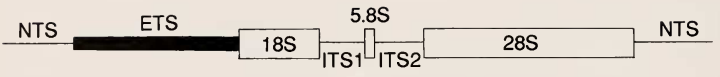

Fig. 1. Generalized diagram illustrating the components of the nuclear rRNA repeat unit (in this case that of vertebrates, after Gerbi 1985), showing the relative positions of the 5.8S, 18S and 28S regions, nontranscribed spacers (NTS), an external transcribed spacer (ETS) and two internal transcribed spacers (ITS).

excellent books by Hillis and Moritz (1990) and Miyamoto and Cracraft (1991); for reviews of molecular techniques and applications for insect systematics see Simon (1991) and Simon et al. (1990, 1991). Ladiges and Martinelli (1990), though focusing on plant systematics, also contains a number of useful general papers on both theoretical and practical aspects of molecular systematics.

CLASSES OF DNA FOR PHYLOGENETIC ANALYSIS

In the last decade, the application of molecular data to systematics has expanded enormously, and comparative DNA sequences have become the preferred data for such investigations (Hillis and Moritz 1990; Miyamoto and Cracraft 1991). Sequence characters from nuclear and extranuclear genomes offer a more or less unlimited supply of diverse characters applicable for analyses at all taxonomic levels, from the population to the Kingdom. Different genes and gene regions exhibit vastly different evolutionary rates, structural or functional constraints, and mutational biases (Nei 1987; Larson and Wilson 1989; Simon et al. 1991), thus it is potentially possible to match specific systematic questions to appropriate genomic regions for analysis. For example, regions of DNA that are evolutionarily conserved, such as sections of rRNA (Gerbi 1985), are useful for resolving early phylogenetic history (Field et al.1988; Lake 1988; Mindell and Honeycutt 1990), whereas regions showing intermediate (Larson and Wilson 1989) or rapid divergence (Brown et al. 1979; Crozier et al. 1989) are useful for evaluating evolutionary events that occur on intermediate (Larson 1991; Cameron 1991) or short (Greenberg et al. 1983) time scales. Nuclear rRNA sequences have been used extensively for the phylogenetic reconstruction of a great diversity of organisms, and more recently, mitochondrial rRNA and protein-coding genes have contributed even more sequence information (Simon et al. 1991). We briefly review each of these classes of DNA.

Nuclear encoded rRNA.— Ribosomes are the sites for cellular protein synthesis and as such their RNA is present in many copies and is abundant compared with cellular mRNA and tRNA. Eukaryote rRNA is composed of two subunits; the smaller subunit has a sediment coefficient of about 18 and is known as 18S rRNA, while the larger subunit comprises three components, viz. 5S, 5.8S and 28S rRNA (Fig. 1). Sequences from the smaller components (5S and 5.8S) are generally inappropriate for phylogenetic analysis because of their size, but the intermediate and larger subunits, particularly 18S rRNA, have been used to examine relationships among a great range of taxa (Johnson and Baverstock 1989; Mindell and Honeycutt 1990; Baverstock and Johnson 1990; Larson 1991). The 18S rRNA is 1700 to 2300 bases long in eukaryotes and as a non-coding region, its insertions and deletions can comprise any number of bases (not limited to multiples of three) because frame shifts do not apply. Furthermore, comparison of sequences indicates that introns are generally absent in rRNA (Baverstock and Johnson 1990).

Some regions of 18S rRNA are moderately variable and have application for lower levels of phylogenetic analysis. However, the more conserved regions have been the focus of many higher-level studies. Indeed, so called 'fossil RNA' exhibits identical sequences (24 bases in length) between organisms as divergent as prokaryotes (e.g., archaebacteria) and eukaryotes (e.g., humans). Generally, 18S rRNA sequences are considered useful for taxa that diverged from 100-1000 Mya (Baverstock and Johnson 1990), while 28S rRNA sequences are useful for divergent times of 60-200 Mya (Larson 1991). Studies to date (Table 1) have examined the relationships between Kingdoms, major prokaryote groupings, protistan phyla, invertebrate phyla, classes of platyhelminths, chordate groups, and vertebrates. Few studies have been published on the phylogeny of insect groups

Table 1. Selected references to studies employing rRNA sequence data for phylogenetic analysis and the corresponding taxa examined.

Reference	Region	Taxon
Nuclear 5S, 5.8S, 18S and 28S rRNA		
Walker 1985	5S & 5.8S	Protistan phyla
Woese et al. 1990	18S	Kingdoms
Fox et al. 1980	18S	Major prokaryote groups
Woese 1987	18S	Major prokaryote groups
Johnson & Baverstock 1989 (review)	18	Sprotistan phyla
Field et al. 1988	18S	Invertebrate phyla
Baverstock et al.1991a	18S	Platyhelminths
Qu et al. 1986	18S	Helminths
Wheeler 1989	18S	Insect orders
Joss et al. 1991	18S	Chordate groups
Baverstock et al. 1991b	18S	Higher vertebrates
Jupe et al. 1988	18S	Algae
Mindell & Honeycutt 1989	18S/28S	Birds
Hedges et al. 1990	18S/28S	Tetrapods
Hamby & Zimmer 1988	18S/26S	Grasses
Zimmer et al. 1989	18S/26S	Flowering plants
Baroin et al. 1988	28S	Unicellular eukaryotes
Vossbrinck & Friedman 1989	28S	Diptera
Larson 1991	28S	Salamander
Hillis & Dixon 1989	28S	Vertebrates
Hillis & Davis 1987	28S	Amphibians
Schmickel et al. 1990	28S	Primates
Sheppard & McPheron 1991	18S/28S	Apidae
Nuclear intergenic spacer (IGS) of rRNA		
Collins et al. 1987	-	*Anopheles* (Diptera)
Beach et al. 1989	-	*Anopheles* (Diptera)
Collins et al. 1989	-	*Anopheles* (Diptera)
Tautz et al. 1987	-	*Drosophila* (Diptera)
Lassner et al. 1987	-	*Triticum* (wheat)
Sheppard & McPheron 1991	ITS1	Apidae
Mitochondrial 12S and 16S rRNA		
Cameron 1991	16S	Apidae
Derr et al., in press	16S	Hymenoptera
Thomas et al. 1989	12S	marsupials
Hixson & Brown 1986	12S	primates
Miyamoto & Boyle 1989	12S/16S	eutherian mammals
Miyamoto et. al. 1989	12S/16S	artiodactyl mammals
Simon et al. 1990	12S	cicadas
Sheppard & McPheron 1991	12S (very divergent)	Apidae

using nuclear encoded rRNA sequences; exceptions include Vossbrinck and Friedman (1989) on cylorrhaphous Diptera , Wheeler (1989) on the Insecta, Sheppard and McPheron (1991) on Apidae. Also, a recent analysis of the blattoid insects has been completed by Vawter (1991 Ph.D. dissertation).

Mitochondrial DNA.— Animal mtDNA is a circular, double-stranded molecule ranging from about 14 kb to 39 kb in length (reviewed in Avise et al. 1987; Moritz et al. 1987). The mtDNA of only one insect, *Drosophila yakuba*, has been completely sequenced (Clary and Wolstenholme 1985). It encodes 13 proteins, 22 tRNAs and two rRNAs, as for most animals. Partial mtDNA sequences are known for other insects, including crickets (Rand and Harrison 1989), mosquitoes (HsuChen and Dubin 1984; HsuChen et al. 1984), cicadas (Simon et al. 1990), and honey bees (Vlasak et al. 1987; Crozier et al. 1989; Garnery et al. 1991; Cameron, unpublished data). For a complete list of published mtDNA sequences see Simon et al. (1991).

In vertebrates, mtDNA has been found to evolve many times faster than single-copy nuclear DNA (scnDNA) (Moritz et al. 1987), in contrast to invertebrates, which exhibit approximately equal rates of change (amino acid or nucleotide substitutions) for both genomes (Vawter and Brown 1986; Powell et al. 1986). In the Hymenoptera, some mtDNA genes are highly conserved (e.g. ND3, Les Willis, unpublished data for *Apis*), while others exhibit rapid rates of divergence (e.g., COII, Crozier et al. 1989; Garnery et al. 1991; rRNA, Cameron, unpublished data; Derr et al., in press). Within both the 12S and 16S rRNA genes, some regions are highly conserved while other regions are rapidly diverging (Cameron, unpublished data; Derr et al., in press). Recent investigations of honey bee mtDNA indicate that it has a significantly greater evolutionary rate than that of *Drosophila* (Crozier 1989); however the causal factors are unknown.

In summary, current knowledge suggests that hymenopteran mtDNA exhibits vastly different rates of evolution, and therefore is useful for phylogenetic inference at many levels. A note of caution, however, when examining relationships below the genus level. The random sorting of polymorphic genes within a species may lead to a lack of congruence between the phylogenetic pattern of the gene (mtDNA) and that of a group of closely related species (Takahata 1989). We discuss below (Examples of Current Research) the usefulness of comparative sequences from the 16S rRNA gene for assessing phylogenetic relationships at three different levels: (1) among species of the genus *Apis*, (2) among tribes of the family Apidae), and (3) among families and superfamilies of Hymenoptera.

Protein-coding genes.— Protein-coding genes (also referred to as structural genes) are transcribed into RNA and then translated into proteins. These have been used less often for phylogenetic analyses, in part because early on, rRNA genes (mito-

chondrial and nuclear) proved useful for many levels of phylogenetic inference (see above). Thus, a large number of rRNA primers have been synthesized, many of which are applied to new studies utilizing sequence data. Fewer primers are available for protein-coding genes. An additional concern is that many nuclear encoded protein-coding genes are parts of multiple-copy divergent gene families, making analysis (and possibly PCR) more complicated. However, the application of mitochondrial protein-coding sequences for phylogenetic studies of Hymenoptera is expanding. In addition to the study of *Apis* relationships by Garnery et al. (1991), full sequences are available for the mitochondrial COI and COII genes of *A. mellifera* L.(Crozier 1989). These have been used to synthesize primers for several current phylogenetic investigations, including another analysis of the genus *Apis* (Les Willis, unpublished data). For a review of current knowledge on mitochondrial protein-coding genes, including primer sequences used for PCR and sequencing, see Simon et al. (1991).

OBTAINING SEQUENCE DATA FOR PHYLOGENETIC ANALYSIS

Although the usefulness of nucleotide sequences for phylogenetic analysis has become widely recognized, the need for technical training in molecular biology, and the time and expense involved in obtaining the data has curtailed widespread use of the technology for systematics. This has been particularly true for Hymenoptera and other insect groups, which are small relative to vertebrates, presenting challenges for extracting DNA in sufficient quantities for sequencing. In addition, many Hymenoptera, especially aculeates, have a hard chitinous exoskeleton which, in contrast to the softbodied *Drosophila*, makes DNA extraction more difficult.

These problems have, in principle, been solved by the revolutionary new development of automated technology for the enzymatic amplification of DNA based on the polymerase chain reaction (PCR) (Saiki et al. 1988; Innis et al. 1990). PCR is a thermocyclic reaction (discussed below) that generates multiple copies of a fragment of DNA relatively quickly and cheaply, eliminating the lengthy procedures of viral or bacterial cloning (Saiki et al. 1985; Mullis et al. 1986; Mullis and Faloona 1987; Cherfas 1990; Innis et al. 1990; Mullis 1990). Although PCR is still in its infancy as a tool for systematics, this method now makes it feasible to obtain large quantities of homologous DNA for direct sequencing from individual insects (Wheeler 1989; Simon et al. 1991; Cameron 1991). Because small amounts of template DNA are sufficient for amplification with PCR, samples no longer must be fresh or frozen, they may be preserved in alcohol or formalin, or even dried (Pääbo 1989; Pääbo 1990; Kocher et al. 1989). Thus, PCR has the capacity to expand phylogenetic investigations to include untapped temporal and geographic coverage of museum specimens.

PCR works with two oligonucleotide primers, which are short pieces of DNA in the range of 18-25 base pairs (bp) in length (see Appendix 1). Each primer is designed to be complementary to one of the two strands of the sample DNA, and together they flank the region to be amplified, which is usually several hundred to several thousand base pairs (kilobases or kb) in length. PCR occurs in three steps, repeated 30-40 times. First, the sample DNA is denatured by heat into its two respective strands. Next, the reaction mixture is cooled to allow the two primers to anneal to their complementary strands. Lastly, in the presence of a thermostable DNA polymerase, such as Taq polymerase (derived from a thermophilic bacterium), the two complementary sample strands are replicated by primer extension, beginning at the primer sites (for figured descriptions see Hillis et al. 1990; Simon et al.1991). The target DNA is therefore replicated exponentially, and within several hours the double-stranded sample has been amplified several millionfold. Single-stranded DNA can be produced by using an excess of one of the primers, a procedure known as asymmetric amplification (Gyllensten and Erlich 1988).

The procedures and protocols for DNA extraction, amplification, purification and sequencing (modified for Hymenoptera) are too extensive to present here and will be published elsewhere (Cameron, unpublished data; Derr et al., in press). However, several recent references provide useful information: Hillis et al. (1990) describe a basic laboratory setup, protocols, and recipes for stock solutions; Innis et al.(1990) provide a thorough description of PCR methodology and its various applications and protocols; Simon et al. (1991) provide up to date information on invertebrate mitochondrial (and other) primer sequences for use with PCR, as well as PCR protocols for use with insect taxa; and Maniatis et al. (1982) is an indis-

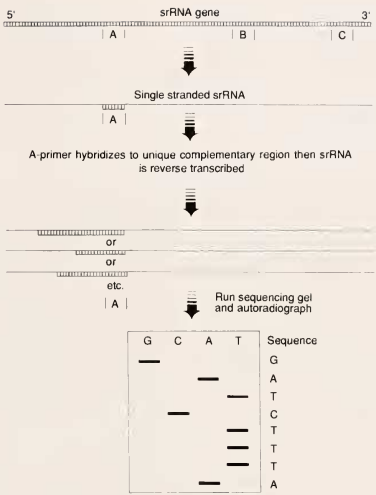

Fig. 2. Diagrammatic representation of the sequencing of small and large subunit rRNA using the reverse transcriptase method. In this case for the 18S rRNA, primer A is added to a bulk RNA extract; it hybridizes to the complementary region and acts to prime the intitiation of DNA sythnesis in a Sanger sequencing reaction (see text for further details) (after Johnson and Baverstock 1989).

pensable reference for many molecular procedures.

Another method of obtaining DNA for sequencing is to isolate the transcribed DNA for a region of RNA (e.g., mRNA) and use reverse transcriptase for sequencing. This method, developed by Qu et al. (1983) and Lane et al.(1985) for obtaining sequences of RNA, provides a relatively quick and easy method of sequencing by using small conserved regions to prime reverse transcription of cellular RNA. In short, sample tissues are treated with guanidine hydrochloride to block RNAase activity. Bulk RNA, consisting primarily of rRNA, is then purified from DNA and protein (see Larson and Wilson 1989; Hillis et al.1990). One of several oligonucleotide primers (e.g., Field et al. 1988; Baverstock et al. 1991a) is then added to the purified RNA and the DNA is then sequenced by chain termination (Sanger et al. 1977) at the same time that it is produced by reverse transcription. The resultant products are then run on a sequencing gel and the sequences read from an autoradiograph (Fig. 2). The entire procedure takes only days instead of the months required for cloning DNA.

ANALYSIS OF SEQUENCE DATA

Sequence Comparison and Alignment.— Probably the most difficult and least understood aspect of the use of sequence data for phylogenetic analysis is sequence alignment (Swofford and Olsen 1990). Phylogenetic analysis of sequence information requires the correct alignment of homologous components between pairs of sequences. One must be careful to distinguish between phylogenetically homologous DNA sequences (orthologs) and multiple, diversified gene copies within single individuals (paralogs) (Fitch 1970, Patterson 1987). In phylogenetic studies, the goal is to compare orthologous sequences from taxa of interest. Superficially, this would seem a rather straightforward task. For example, each nucleotide position can be viewed as a 'character' with only a limited number of 'states' possible at each position (i.e., for DNA sequences, 'A', 'C', 'G', 'T', or a gap mutation

Suborder	Superfamily	Genus	Nucleotide Sequence
Symphyta	Siricoidea	*Tremex*	AA-ATATAAATTAAATTCT-
Apocrita	Vespoidea	*Polistes*	AA-AACATTTTTAAATTCT-
Apocrita	Ichneumonoidea	*Xanthopimpla*	ATTAATA-AATTAAA-GCTC

Fig. 3. One possible sequence alignment from a small segment of the large ribosomal subunit (16S rRNA) of three representative hymenopteran taxa. This region corresponds to positions 13,205 to 13,225 of the published *Drosophila* sequence (Clary and Wolstenholme 1985) (data are from Derr et al., in press). From a total of 20 nucleotide positions from three taxa there are six inferred gap mutations, only six G or C bases, and 48 A or T bases. There are 10 possible base substitutions of which nine are transversions (involving A/T, A/C or G/T bases) and only one transition (T/C).

'-'). In practice, however, alignment of multiple nucleotide sequences can involve a number of complicating factors. For example, as discussed by Swofford and Olsen (1990), in addition to requiring the use of orthologous sequences, phylogenetic analysis of sequence data requires that one make the assumption that all nucleotides observed at a given position are traceable to a common ancestor. Historical events such as insertions, deletions, duplications, rearrangements, and multiple nucleotide substitutions all combine either to cloud the evolutionary history of some nucleotide positions or make non-homologous positions indistinguishable.

Determination of sequence homology and alignment usually presents few ambiguities when working with protein-coding gene sequences, particularly scnDNA. Landmark features along these sequences such as codons (three adjacent bases specifying an amino acid), intron/exon junction consensus sequences, various start and stop signals, and other DNA/protein conserved binding sites provide clues that make alignment of these sequences straightforward. These landmark features are especially useful for alignment when very distantly related taxa are compared. Moreover, positions within each codon tend to evolve at different rates, with third position changes being most frequent, first position changes being highly conserved, and second position changes somewhere in between. Therefore, the reading frames in protein coding sequences provide an inherent structure useful in alignment.

Nonprotein-coding regions, such as rRNA, tRNA and other non-translated sequences are potentially more difficult to align with distantly related taxa, due in part to the lack of these landmark features. In addition, these sequences usually are characterized by nucleotide base insertion and deletion events, presumably because there are no selective constraints to maintain reading frames that code for specific amino acids (Mindell 1991). In practice, however, the alignment of most nuclear-encoded rRNA sequences by eye does not seem to have posed significant problems because of the relatively small number of insertions/deletions and the general conservative nature of the subunits. Some regions of mitochondrial 16S rDNA may pose alignment problems in Hymenoptera because they exhibit an unusually high frequency of A's and T's relative to G's and C's (Cameron, unpublished data; Derr et al. in press). Consequently, nucleotide substitutions in these areas may be characterized by a high proportion of transversions (purine (A/G) to pyrimidine (T/C) substitutions, or the reverse) as opposed to the bias toward transitions (purine - purine or pyrimidine - pyrimidine) commonly observed in vertebrate mitochondrial genomes (Hixson and Brown 1986; Thomas and Beckenbach 1989). This becomes important when using computer alignment schemes (discussed below), which generally assign higher penalties to transversion substitutions. Also, considerable length polymorphism is evident in these AT-rich regions; large insertions and deletions (often greater than 10 base pairs in length) can further complicate alignment because of uncertain homology among the bases. It is best to exclude these hypervariable regions from the analysis. Sequences from the 16S rRNA region are depicted for three hymenopteran taxa in Fig. 3, taken from the study of Derr et al. (in press). These provide examples of insertion/deletion events, strong A/T base compositional bias, and a correspondingly high rate of transversion over transition substitutions. As a consequence of these factors, most of the

difficulties encountered in aligning nucleotide sequences involve nonprotein-coding regions.

Algorithms designed to determine the optimal alignment between two sequences have been available for some time (Needleman and Wunsch 1970; Sankoff 1972; Sellers 1974; Waterman et al. 1976). These programs attempt to maximize the number of matches or to minimize the number of substitutions, insertions, or deletions required to make two sequences equivalent (Mindell 1991). However, extending this approach to more than two sequences with no *a priori* regard to their phylogenetic relationships has been deemed inappropriate for reconstructing phylogenies (Hein 1989, 1990b; Feng and Doolittle 1987, 1990). These authors contend that in multiple alignments the initial choice of sequences for pairwise comparison can bias the final alignment, result in an excess of inferred gap events, and even affect phylogenetic results (Lake 1991). Therefore, multiple sequence alignment, at least in principle must fall under the same constraints used to infer phylogeny, in this case, global parsimony or minimizing the overall number of substitution and gap events. Alignment of sequences could be considered as part of phylogeny inference, rather than as an independent analysis (Sankoff et al. 1973). Moreover, phylogenetic congruence with other independent data sets offers a means of choosing among equally parsimonious alignments (Hillis et al. 1990).

Alignment of nucleotide sequences may be accomplished by hand and computer. Several 'progressive alignment' computer programs are currently available (e.g., Higgins and Sharp 1988; Higgins et al. 1992; Hein 1989). These programs generally proceed by: (1) calculating an initial similarity value for each pairwise comparison of sequences; (2) constructing a dendrogram by cluster analysis using the matrix of these values; and (3) aligning the sequences according to the branching order in the dendrogram. Alignment scores are calculated by assigning positive or negative values to matches and mismatches and by imposing penalties for both the insertion of gaps and for each additional change within a gap. In most cases the user may assign a numerical value for each of these penalties. Aligned sequences may then be analyzed phylogenetically using any of the currently available parsimony-based computer packages. Measures of homoplasy in the results can also be used to discriminate among various sequence alignments. In practice, results of computer alignment procedures should always be compared with those obtained by hand, and we have found (Cameron, unpublished data; Derr et al., in press) that final computer alignments can be fine-tuned by visual inspection.

At present our understanding of the complexities of sequence comparison and analysis is still incomplete but developing rapidly. Our intent here has been to highlight the problems inherent in multiple sequence alignment as it relates to phylogenetic reconstruction, and to indicate some methods available for their solution. In general, sequence alignment is straightforward when dealing with single-copy protein-coding sequences; with non-protein coding sequences the researcher should be aware that in areas with few conserved landmark features, sequence alignment can present a number of experimental challenges. Fortunately, as the field of molecular systematics continues to evolve and as more comparative sequence data becomes available, these challenges will be met by the development of increasingly useful and realistic computer alignment algorithms. For information regarding the algorithms discussed here, refer to the work of Sellers (1974), Smith et al. (1981, 1985), Feng and Doolittle (1987, 1990). For further information on sequence alignment, homology, and weighting schemes see Mindell (1991); for general reviews see Bell and Marr (1989), Doolittle (1990), Hillis et al. (1990), Hein (1989), and Watermann et al. (1991).

Phylogenetic Analysis of Sequence Data.— Many methods have been proposed for reconstructing phylogenetic relationships with DNA sequence data for three or more taxa, and we do not propose to review them all here. Swofford and Olsen (1990) and Felsenstein (1988) provide excellent recent reviews of distance, maximum likelihood, and parsimony methods, and comment on both the logical foundations of various approaches and the 'nuts and bolts' issues of actually getting the job done. Of the various approaches currently in use, we favor a simple parsimony model for reasons of simplicity and clarity, both in analysis and in the interpretation of results. With correctly aligned sequences, parsimony analysis is relatively straightforward. Each nucleotide position is treated as an independent, unweighted character with four possible states: adenine or guanine (purines) or cytosine or thymine (pyrimidines). The simplest approach is to treat a substitution from one base to any other as equally likely ('Fitch parsimony', Fitch

1971) and this can be accommodated by treating the characters as 'unordered' or 'non-additive' (terminology differs between programs). However, because of structural constraints on the DNA molecule itself, a bias toward transitions (purine-purine or pyrimidine-pyrimidine) and against transversions (purine-pyrimidine or the reverse) has often been noted (e.g. Li et al. 1984, Hixson and Brown 1986). Some programs offer tools to accommodate differential weighting of some character state changes over others. For example, the 'Step Matrix' function in PAUP (Swofford, 1990) can be used to assign any integer weight for changes between any two character states. Of course, the problem is to determine what weights to assign. In highly AT-rich sequences such as are found in many Hymenoptera, many changes necessarily will be from A to T or the reverse (transversions), and there may not be a bias towards transitions. It is possible, at least in theory, to examine empirically the base composition of sequences and to derive from these the expected probabilities for the different categories of substitution. Swofford and Olsen (1990) make a sensible suggestion: by giving only slightly lower weight to transversions, the weights will come into play only in choosing between essentially equally parsimonious solutions, and transitions will then be given the edge.

A strategy to reduce the effects of homoplasy with sequence data from protein-coding genes is to eliminate nucleotides in the third position of each codon, or to give them a lower weight. This is based on the redundancy of the DNA code; that is, in most cases the first two positions of the code are sufficient to specify an amino acid and the third may be redundant information. As a result, substitutions in third positions may accumulate more rapidly than in the other positions. If more than one substitution has taken place, the position is no longer informative. Although this approach is usually limited to protein-coding genes, in an analogous fashion, if the secondary structure of non-protein-coding sequences is known, regions shown to be undergoing compensating substitutions can be eliminated (Wheeler and Honeycutt 1988), or preferably, given an appropriate lower weight (Vawter, 1991).

For analysis of small data sets, any up to date computer algorithm for parsimony analysis will suffice but we recommend using a recent version of one of the readily available algorithms such as PAUP (Swofford 1990) or Hennig86 (Farris 1988). For studies with fewer than 15-20 terminal taxa, one of the exact methods can be used (branch and bound, exhaustive search, or implicit enumeration), and one can be confident that the most pasimonious tree or trees have been found. For larger datasets or those with relatively high levels of homoplasy, heuristic search procedures such as branch-swapping will be required. In such cases, it is important to try many different addition sequences and search procedures, until one's patience has literally been exhausted, because it is often difficult to escape local optima in which the algorithms become trapped, or to find all of the different groups (or 'islands') of equally parsimonious solutions (Maddison 1991).

A problem that is more or less unique to sequence data is how to handle insertion and deletion events, for example, as inferred by alignment procedures. A conservative approach is to treat gaps in sequences as missing data. In this case they will have no effect on tree length or character state optimization. However, insertions and deletions may represent real phylogenetic events and this approach ignores their potential contribution to phylogenetic reconstruction. An alternative is to treat insertions and deletions as separate characters, but if they vary in length, one will encounter problems in establishing their homology and the transformation series among them. One's choice of approach should be governed directly by the data.

Outgroup Selection.— Outgroups may be used to determine character polarity or to root unrooted trees following a parsimony analysis (Watrous and Wheeler 1981; Donoghue and Cantino 1984; Maddison, Donoghue, and Maddison 1984). For Hymenoptera, selection of an outgroup for taxa at the rank of subfamily or above is often problematical. For example, although the Symphyta are perhaps best thought of as a basal paraphyletic group within the Hymenoptera, there are several competing hypotheses of relationships among symphytan groups (Ross 1937; Königsmann 1977; Rasnitsyn 1980, 1988; Gibson and Goulet 1988). These alternative hypotheses affect both the choice of an outgroup for the remaining Hymenoptera (the Apocrita) as well as hypotheses of character state evolution within various symphytan lineages. Within the Apocrita, relationships among the nonaculeates are particularly problematical. Recent suggestions (Rasnitsyn 1988; Mason, unpublished data) that the Aculeata are the sister group to the Ichneumonoidea would have a significant impact

on character polarities within those groups, but this hypothesis remains relatively untested. Similar problems are apparent for larger groups throughout the order. Indeed, at the ordinal level, there is virtually no agreement on the appropriate sister group to the Hymenoptera as a whole. Until higher level relationships among the Insecta are better known, the choice of an appropriate outgroup will continue to be a problem for studies of phylogenetic analyses within Hymenoptera, regardless of the type of evidence used.

Why might the choice of outgroup be critical when using molecular data for phylogeny? Wheeler (1990) has recently discussed some of the problems posed by distant or uncertain outgroups when using molecular data. If distantly related taxa are used as outgroups, the probability that sequence similarity is due to random identity increases, and the chance that any one character is phylogenetically informative consequently decreases. If outgroup taxa are sufficiently divergent, polarization of characters essentially becomes random. In essence, results become phenetic, rather than phylogenetic.

How can this affect results of parsimony analyses? If sequences are too divergent, an ingroup may not be resolved as monophyletic relative to multiple outgroups. We encountered this problem when using two Diptera, *Aedes* and *Drosophila*, as outgroups to Hymenoptera (Derr et al., in press). The two dipterans were nearly as divergent from each other as they were from some of the Hymenoptera, resulting in instability at the base of the tree. In fact, in this case Hymenoptera could not be resolved as monophyletic relative to Diptera, clearly an unsatisfactory result.

Assessing the Reliability of Results.— Bootstrap methods in combination with parsimony procedures have become popular in recent years as a way to assess the degree of support for a particular phylogenetic clade (Felsenstein 1985, 1988). Bootstrapping involves random sampling with replacement from a set of characters until a new character set is formed, equal in number to the original set. From this new character set, another maximum parsimony tree is estimated. The procedure is repeated many times (e.g. 100-10,000) and a distribution of solutions is obtained. Several assumptions underly bootstrap analysis (Felsenstein 1985, 1988). One is that nucleotides evolve entirely independently of one another, that DNA initially consists of unlinked sequences of nucleotides that change at random throughout the molecule. Another is that nucleotides are identically distributed for all taxa. The first assumption may be violated with hymenopteran mtDNA. Our investigations (discussed below) indicate that hymenopteran mtDNA is highly AT-rich and that A-T transversions are far more likely than other types of transversional substitutions. This is a clear violation of the equal probability assumption, which predicts that only 1/4 of all transversions should be A-T transversions. Another violation of the independence assumption arises with sequence data if secondary structural constraints in the molecule result in compensating substitutions (Wheeler and Honeycutt 1988; Simon et al. 1990), such as Vawter (1991) found for a relatively small number of bases in the stem region of insect rRNA. However, with bootstrap one can take these biases into account (just as with parsimony analysis) by applying, for example, less weight to A-T transversions or to sequences with known compensating substitutions. The second assumption, that characters are identically distributed, poses a difficulty if sequences are selected from different regions of the genome with different distributions. Also, mixing morphological and molecular data in a single bootstrap analysis would violate this assumption if the two character sets reflect different distributions (e.g., continuous and discrete; normal and Poisson). Bootstrap percentages are often interpreted as confidence intervals associated with particular topologies. However, this is appropriate only for testing the validity of a single lineage that has been identified in advance of the analysis (Swofford and Olsen 1990) and if the above assumptions are met. Violation of these assumptions severely reduces the accuracy of the reported confidence intervals (Sanderson, 1989). Even though the assumptions of the bootstrap may be restrictive, it is nonetheless a valuable heuristic method for testing the robustness of results from parsimony analyses. For example, the appearance of a particular group or clade in all or most (e.g. 85-95%) of the replicates may be used as an index of support for its monophyly (Swofford and Olsen 1990).

One may also wish to know whether a given set of characters support one tree topology significantly more strongly than another topology under the assumption of parsimony. Templeton's paired comparisons test compares two trees in this fashion. This test is an application of Wilcoxon's nonparametric signed-ranks test, using Templeton's criteria for nucleotide sequence data (Templeton

1983). The scoring procedure involves counting the number of substitutions at each informative site for two given trees and applying the Wilcoxon test to the hypothesis that the total number of substitutions is equal for the two trees. Also, additional information can be incorporated into the scoring procedure. If, for example, it is known that transversions are more common than transitions in a given region of DNA, one might choose to give more weight to transitions by assigning them a higher score than transversions (e.g., transitions = 2, transversions = 1). Templeton's test is conservative, thus it is difficult to reject the null hypothesis without large differences in the number of substitutions between the two trees. This is one of the strengths of the procedure.

Felsenstein has developed a test that uses similar data to investigate relationships among sets of three taxa, following the statistical approach of Cavender (1981). Cavender pioneered methods for applying confidence intervals to phylogenies based on parsimony, and Felsenstein (1985) later modified Cavender's methods to include sequence data. For a group of three taxa (rooted with an outgroup), there are three possible alternative tree topologies. Two statistics are used to evaluate whether the most parsimonious topology is significantly better than the other two at the 95% confidence level: *S* is the number of additional steps that a tree must have to be significantly worse than the most parsimonious tree; *C* is the number of phylogenetically informative characters that must support a tree topology for it to be significantly better than the others (Felsenstein 1985). Like Templeton's test, this test is conservative; a tree topology that differs by only a few steps, or is supported by only a few more characters will not be significantly different. Felsenstein's test assumes a molecular clock (i.e., that the number of changes in a lineage is roughly proportional to the amount of time since its divergence), a controversial assumption which has only just begun to be examined in insects (Crozier et al. 1989).

One final caveat: if data are homoplastic, multiple models of character state change may be possible on a given minimum-length tree topology. The simplest example is a case in which a parallelism or a reversal is equally parsimonious, and either may be used to explain the data. Parsimony programs contain a number of tools, known as 'optimization' methods, to assist in modeling character state change on cladograms. We caution against the use of any one criterion, for example, minimizing parallelisms or reversals. In principle, the best approach is to determine all of the possible alternative models of character state change for each equally parsimonious tree topology. In practice, this is feasible only if relatively few characters are involved; with sequence data the alternatives are likely to be numerous and complex. A workable alternative is the use of tree diagnostics, which show the minimum and maximum number of steps possible in each interval on the tree under all possible models of character state change.

EXAMPLES OF CURRENT RESEARCH

Tribal Phylogeny of the Family Apidae.— The focus of this investigation was to examine the usefulness of DNA sequence data for resolving phylogenetic relationships among tribes of the family Apidae. Sequences from the mitochondrial 16S rRNA gene were compared in 15 exemplars representing the four apid tribes (Cameron, 1991). The exemplar approach was justified on the basis that the tribes (considered as subfamilies by Michener 1990) have been recognized as monophyletic groups. The use of several taxa (as many as is practicable for the study) to represent each clade is important for several reasons. First, the use of multiple taxa will help to resolve the degree of sequence variation exhibited within a given region (e.g., variation among species within a tribe or variation among tribes), hence assist in the selection of regions appropriate for a given level of inference. Second, using multiple exemplars from each clade should help to eliminate random error or potential biases that could affect the evaluation of alternative phylogenies. At least two individuals of each species were sequenced as a check against sequencing errors and potential intra-specific variation. Sequences were obtained from fresh, frozen, and ethanol-preserved tissue. The outgroups for the analysis were selected from the subfamily Xylocopinae (family Anthophoridae), considered to be monophyletic and the closest relatives of Apidae (Sakagami and Michener 1987). Two outgroups were selected from two different xylocopine tribes (Xylocopini and Allodapini).

Between 500 and 600 bp were sequenced from the 3' end of the 16S rRNA for all 17 taxa. Sequencing was accomplished using two primers (Fig. 4; Appendix 1) designed to optimize the match between published sequences from the 16S mitochondrial

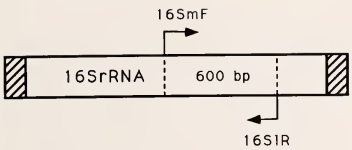

Fig. 4. A representation of the mitochondrial 16S (large subunit) rRNA gene, flanked by transfer RNAs (hatching). The two outside arrows correspond to the position and direction of extension of the oligonucleotide primers used in PCR and sequencing reactions (Cameron, 1991, unpublished data). Dotted lines circumscribe the approximate 600 bp region of the gene that was amplified with PCR.

rRNA of the honey bee *Apis mellifera* L. (Vlasak et al. 1987) and partial sequences obtained for other apid taxa with the use of 'universal' primers (Kocher et al. 1989; John Patton, unpublished data). From the total number of nucleotides sequenced, 116 were informative in the sense that at least two ingroup taxa shared substitutions at those sites. A gap was considered as a fifth character, which did not give undue weight to deletions as gaps were rare among the informative sites. Length polymorphisms were evident in the 16S rRNA of every taxon, but these were not included as characters. Transition and transversion substitutions were treated with equal weight in the analysis. The sequences were aligned by hand and checked by computer alignment using the Treealign Computer Program (Hein 1990a). The issues of length polymorphisms, character weighting, and alignment are treated in detail above (see Phylogenetic Analysis of Sequence Data).

The 116 informative sites from the 15 ingroup taxa and one of the outgroups, *Xylocopa virginica* (L.), were analyzed using maximum parsimony techniques implemented in PAUP (Version 3.0L, Swofford 1990). Maximum likelihood (Felsenstein 1981) and bootstrap analyses (Felsenstein 1985) were implemented as heuristic methods to test for the reliability of the results based on maximum parsimony. Two equally parsimonious trees were produced (Figs 5A, 5B). In tree A, Apini + Euglossini comprise one clade and Bombini + Meliponini comprise a second clade. In tree B, the Bombini + Meliponini clade is retained, with Euglossini as its sister group. The results are consistent with monophyly of each of the currently recognized tribes, except Bombini, which appears to be paraphyletic with respect to Meliponini (trees indicating the monophyly of Bombini were only two steps longer). Both bootstrap and maximum likelihood analyses strongly supported the Bombini + Meliponini clade. To test for effects of the choice of outgroup, an additional outgroup (Allodapini: *Exoneura*) was included in a separate analysis. This resulted in two maximum parsimony trees, each with the same tribal topology as tree A (Fig. 5). Future work should include additional analyses of more distantly related outgroup taxa from the Anthoporidae.

The sequence information obtained from the 16S region had some interesting characteristics, including a higher proportion (> 80%) of A and T bases and a correspondingly high number of transversion-substitutions. Length polymorphisms almost exclusively comprised strings of A's and T's. The occurrence of large insertions and deletions resulted in the exclusion of sections of the sequences from the analysis because of questionable alignment. Nonetheless, this region was sufficiently conserved overall to be useful for resolving relationships at the tribal level. The instability of the apini branch can probably be resolved by including sequence information from additional representatives of the Euglossini. Because of space considerations, the aligned sequences and information regarding percent sequence divergence,

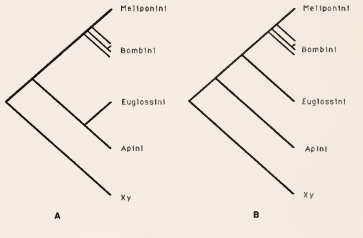

Fig. 5. The two most parsimonious trees (A and B) for the tribes of Apidae, inferred using the branch and bound method implemented in PAUP (from comparisons of nucleotide sequences of mtDNA [16S rRNA] from 16 taxa). The trees are simplified to show only the tribal topology. The outgroup is *Xylocopa virginica* (Anthophoridae). Tree length for analyses of 116 informative sites in 16 taxa was 304 steps, resulting in a consistency index of 0.533.

base composition, and base distribution have been omitted and will be presented elsewhere (Cameron, unpublished data).

Species Phylogeny for the Genus Apis.— The same sequences from the 16S mitochondrial rRNA subunit (500-600 bp) discussed above were used in a separate analysis of five species of the genus *Apis*. These included *A. mellifera*, *A. cerana* F., *A. koschevnikovi* Buttel-Reepen, *A. dorsata* F., and *A. florea* F. One or more exemplars were selected from each of the three remaining apid tribes (Meliponini, Bombini *s.s.*, and Euglossini) and Xylocopinae (*X. virginica*) to serve as outgroups. This study represents a case in which comparative sequences for a given region are useful for two different levels of analysis. From the original data set (above) there were 36 informative sites within *Apis*. Maximum parsimony trees, based only on the informative sites, were estimated in separate analyses using each of the outgroups. Two equally parsimonious ingroup trees were produced: (Figs 6A, 6B). Tree A is concordant with recent analyses based on morphology (Alexander 1991) and comparative sequences from the mitochondrial subunit II of the cytochrome-oxydase gene (COII) (Garnery et al. 1991). A well-corroborated pattern of this nature, utilizing three independent data sets, is highly desirable for two reasons: (1) it suggests a high level of reliability in the phylogenetic pattern, and (2) offers strong support for the acceptance of hypothesis A over hypothesis B (Fig. 6). A complete discussion of these results will appear elsewhere.

Relationships Among the Higher Levels of Hymenoptera: mtDNA.— The focus of this study was to examine the phylogenetic utility and the degree of resolution provided for various hierarchical levels within Hymenoptera by nucleotide sequence information from the 16s rRNA region of the mitochondrial genome (Derr et al., in press). Representative DNA sequences from two members of the suborder Symphyta (superfamilies Siricoidea and Tenthredinoidea) and seven from the suborder Apocrita (superfamilies Ichneu-monoidea, Chalcidoidea and Vespoidea) were examined and compared. In addition, published 16s rRNA sequences from *Aedes* (HsuChen et al. 1984) and *Apis* (Vlasak et al. 1987) were included in the analysis. Multiple individuals and clones were sequenced from each taxon. We were able to obtain usable sequences from specimens killed and preserved in 70% ethanol. Sequences from smaller species (*Aphytis*, Aphelinidae) were obtained from pro-

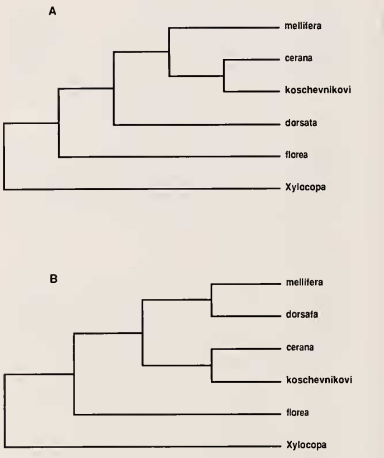

Fig. 6. The two most parsimonious trees (A and B) for *Apis*, inferred using the exhaustive search method implemented in PAUP from comparisons of nucleotide sequences of mtDNA (16S rRNA) from 6 taxa. The outgroup is *Xylocopa virginica*. Tree length for analyses of 6 taxa was 118 steps for 36 informative sites, resulting in a consistency index of 0.643.

geny of single females (isolines). Details regarding DNA isolation, PCR, cloning and sequencing will be provided elsewhere (Derr et al. in press).

Following computer-assisted alignment of the sequences, a total of 573 nucleotide positions was reported with 287 variable in two or more taxa. Each of these sequences was characterized by numerous insertion/deletion events and a bias for A and T bases (cf. Fig. 3). Percent A and T ranged from 0.533 to 0.794, with sequences from members of Ichneumonoidea and Chalcidoidea displaying significantly lower A and T averages. Moreover, sequences from both of these superfamilies also exhibited significantly more strand asymmetry, with an unequal number of purines (A and G) or pyrimidines (T and C) on each DNA strand.

Pairwise comparison among all taxa revealed percent sequence differences ranging from a low of < 2.5% to a high of slightly over 50%. These se-

quences were analyzed using maximum parsimony and bootstrap procedures available in PAUP (Swofford 1990). This analysis, which included rooting with 16s rRNA sequence from the dipteran *Aedes*, resulted in a single most parsimonious tree (Derr et al. in press). A bootstrap consensus tree derived from 100 replications had an identical topology. Two major groups of hymenopteran taxa emerged; the first included the symphytans and the aculeates, the second comprised the parasitic Hymenoptera. However, examination of less parsimonious trees revealed another solution only two steps longer (out of 703 total steps) in which the Symphyta form a basal grade to a monophyletic apocritan clade (aculeates plus parasitic Hymenoptera). All internodes were well supported in the bootstrap analysis with the notable exception of those leading to the Symphyta and the aculeates.

An additional analysis, using one of the two symphytans as an outgroup to the apocritan taxa, also resulted in a sister group relationship between the aculeates and the parasitics. This confirmed the instability at the base of the tree and suggested that a high level of sequence divergence precludes using this region for resolving relationships at the subordinal level. Nevertheless, these results support both the aculeates and at least these parasitic Hymenoptera as distinct monophyletic groups, and provided baseline information regarding the amount and type of nucleotide sequence information available from this region. Interestingly, the sister group relationship between Ichneumonoidea and Chalcidoidea is probably the most strongly supported result to emerge from the analysis. Among the parasitics, the three representatives of the Ichneumonoidea form a monophyletic group. However, sequence divergence among the ichneumonoids was low, providing little resolution among the terminal taxa. Conversely, the two chalcidoid sequences examined, both from the genus *Aphytis*, clearly represented a monophyletic group and they are very divergent from one another, suggesting that this region may have considerable utility at the species level. Baseline information of this type allows subsequent investigations to focus on areas of the genome most likely to produce phylogenetically useful information.

Relationships Among the Higher Levels of Hymenoptera: Nuclear rRNA.— An exploratory study to examine the the usefulness of partial sequences of the small-subunit rRNA for higher-level phylogenetic applications within the Hymenoptera has recently been completed (Austin et al. unpublished data). Although the final results of our investigation are not yet available (and will be published elsewhere), some points can be made that should prove useful to workers who are interested in the molecular systematics of Hymenoptera.

We wished to examine three hypotheses: the paraphyly of the Symphyta, the basal position of the Stephanidae to the rest of the Apocrita, and the sister-group relationship between the Ichneumonoidea and Aculeata (see Whitfield this issue for more information and references to these hypotheses). Initial trials were made with five divergent taxa (*A. mellifera*, *Perga dorsalis* Leach, *Sirex noctilio* F., *Megarhyssa nortoni* (Cresson) and *Ibalia leucospoides* Hochenwarth). Multiple species from some of these five lineages were examined to check the reliability of these data and to test the various methods of analysis against confirmed monophyletic groups. Overall for the ingroup, we collected sequence information from three ichneumonoids (two ichneumonids and a braconid), two pergid sawflies and two aculeates (*A. mellifera* and *Myrmecia* sp.). Because the sister group to the Hymenoptera is unknown, we employed multiple outgroups: *Drosophila*, *Artemia* (published sequences, Dams et al. 1988), and two species of water beetle (newly sequenced as part of another study).

Results obtained using three commercially available universal primers for the 18S subunit rRNA (A, B and C, Field et al. 1988; Baverstock et al. 1991a) revealed a mean sequence divergence of about 5% among the taxa. Two other primers (D and E, Baverstock et al. 1991a, 1991b), reportedly specific to more variable regions of the 18S subunit (Baverstock et al. 1991b), yielded sequences with 3.3% to 17% divergence. These regions proved too conservative to test the above hypotheses. Additional sequence information collected from six other species basically confirms the high degree of conservation within the small-subunit ribosomal RNA, a result which is consistent with those of Sheppard and McPheron (1991) for the Apidae.

It is our opinion that sequences from the large-subunit (28S) rRNA, which have been useful in a preliminary investigation of higher level apid relationships (Sheppard and McPheron 1991), combined with mtDNA and nuclear DNA sequences obtained with PCR technology, will be most fruitful for examining hypotheses of relationships among suborders and families of the Hymenoptera.

ACKNOWLEDGMENTS

Fruitful discussions with P.R. Baverstock, N. Cramp, S.K. Davis, A.M. Johnson, and J.B. Whitfield substantially added to the breadth of the manuscript. Mike Rose provided isolines of *Aphytis* species. Financial support to A.D.A. was provided by the Australian Research Council, to S.A.C. by the National Institutes of Health (grant GM31571), and to J.B.W. by Texas Advanced Research Program grant 999902-044. We wish to thank Emma Cabot for retyping an early draft of the manuscript.

LITERATURE CITED

Alexander, B.A. 1991. Phylogenetic analysis of the genus *Apis* (Hymenoptera: Apidae). *Annals of the Entomological Society of America* 84: 137-149.

Avise, J.C., J. Arnold, R.M. Ball, E. Bermingham, T. Lamb, J.E. Neigel, C.A. Reeb and N.C. Saunders 1987. Intraspecific phylogeography: the mitochondrial DNA bridge between population genetics and systematics. *Annual Review of Ecology and Systematics* 8: 489-522.

Baroin, A. R. Perasso, L.H.Qu, G. Brugerolle, J.P. Bachellerie and A. Adoutte 1988. Partial phylogeny of the unicellular eukyotes based on rapid sequencing of a portion of 28 S ribosomal RNA. *Proceedings of the National Academy of Sciences, USA* 85: 3474-3478.

Baverstock, P.R., R. Fielke, A.M. Johnson, R.A. Brays and I. Beveridge. 1991a. Conflicting phylogenetic hypotheses for the parasitic platyhelminths tested by partial sequencing of 18S ribosomal RNA. *International Journal for Parasitology* 21: 329-339.

Baverstock, P.R., N. Cramp, A.M. Johnson and S. Donnellan 1991b. Evolution of the small subunit ribosomal RNA in higher vertebrates. *Australian Journal of Zoology* 39: 387-401.

Baverstock, P.R. and A.M. Johnson 1990. Ribosomal RNA nucleotide sequence: a comparision of newer methods used for its determination, and its use in phylogenetic analysis, pp. 101-110. *In* Ladiges, P.Y. and L.W. Martinelli, eds., *Plant Systematics in the Age of Molecular Biology*. CSIRO, Melbourne. (Also reprinted in *Australian Systematic Botany* 3: 101-110.)

Beach, R.F., D. Mills and F.H. Collins 1989. Structure of ribosomal DNA in *Anopheles albimanus* (Diptera: Culicidae). *Annals of the Entomological Society of America* 82: 641-648.

Bell, G.I. and T.G. Marr, eds. 1989. *Computers and DNA: The Proceedings of the Interface Between Computation Science and Nucleic Acid Sequencing Workshop*. Santa Fe Institute Studies in the Series of Complexity. Vol. 7. Addison-Wesley, Redwood City, CA.

Brown, W.M., M. George and A.C.Wilson 1979. Rapid evolution of animal mitochondrial DNA. *Proceedings of the National Academy of Sciences, USA* 76: 1967-1971.

Cameron, S.A. 1991. A new tribal phylogeny of the Apidae inferred from mitochondrial DNA sequences, pp. 71-87. *In* Smith, D.R. ed., *Diversity in the Genus Apis*. Westview Press, Boulder.

Cherfas, J. 1990. Genes unlimited. *New Scientist* 126 (No.1712): 17-21.

Clary, D.O. and D.R. Wolstenholme 1985. The mitochondrial DNA molecule of *Drosophila yakuba*: nucleotide sequence, gene organization, and genetic code. *Journal of Molecular Evolution* 22: 252-271.

Collins, F.H., M.A. Mendez, M.O. Rasmussen, P.C. Mehaffey, N.Y. Besansky and V. Finnert 1987. A ribosomal RNA gene probe differentiates member species of the *Anopheles gambiae* complex. *American Journal of Tropical Medicine and Hygiene* 37: 37-41.

Collins, F.H., S.M. Paskewitz and V. Finnerty 1989. Ribosomal RNA genes of the *Anopheles gambiae* complex. pp. 1-28. *In* Harris, K.F. ed., *Advances in Disease Vector Research*, Vol. 6. Springer-Verlag, New York.

Crozier, R.H., Y.C. Crozier and A.G. Mackinlay 1989. The CO-I and CO-II region of honey bee mitochondrial DNA: evidence for variation in insect mitochondrial evolutionary rates. *Molecular Biology and Evolution* 6: 399-411.

Dams, E., L. Hendricks, Y. Van der Peer, J.M. Neefs, G. Smits, I. Vandnbempt and R. De Wachter 1988. Compilation of small subunit RNA sequences. *Nucleic Acids Research* 16: r87-r173.

Derr, J. N., S.K. Davis, J. B. Woolley, and R. A. Wharton. In press. Variation and the phylogenetic utility of the large ribosomal subunit of mitochondrial DNA from the insect order Hymenoptera. *Molecular Phylogenetics and Evolution*.

Donoghue, M.J. and P.D. Cantino 1984. The logic and limitations of the outgroup substitution approach to cladistic analysis. *Systematic Botany* 9: 192-202.

Doolittle, R.F., ed. 1990. *Molecular Evolution: Computer Analysis of Protein and Nucleic Acid Sequences*. Methods in Enzymology, Vol. 183. Academic Press, New York.

Farris, J.S. 1988. Hennig86. Phylogeny inference package available from the author (41 Admiral Street, Port Jefferson Station, New York, NY 11776, U.S.A.).

Felsenstein, J. 1981. Evolutionary trees from DNA sequences: a maximum likelihood approach. *Journal of Molecular Evolution* 17: 368-376.

Felsenstein, J. 1985. Confidence limits on phylogenies: An approach using the bootstrap. *Evolution* 39: 783-791.

Felsenstein, J. 1988. Phylogenies from molecular sequences: inference and reliability. *Annual Review of Genetics* 22: 521-565.

Feng, D. and R.F. Doolittle 1987. Progressive sequence alignment as a prerequisite to correct phylogenetic trees. *Journal of Molecular Evolution* 25: 351-360

Feng, D. and R.F. Doolittle 1990. Progressive alignment and phylogenetic tree construction of protein sequences, pp. 375-387. *In* Doolittle, R.F. ed., *Molecular Evolution: Computer Analysis of Protein and Nucleic Acid Sequences*. Methods in Enzymology, Vol. 183. Academic Press, New York.

Field, K.G., G.J. Olsen, D.J. Lane, S.J. Giovannoni, M.T. Ghiselin, E.C. Raff, N.R. Pace, and R.A. Raff 1988. Molecular phylogeny of the animal kingdom. *Science* 239: 748-753.

Fitch, W.M. 1970. Distinguishing homologous from analogous proteins. *Systematic Zoology* 19: 99-113.

Fitch, W.M. 1971. Towards defining the course of evolution: minimal change for a specific tree topology. *Systematic Zoology* 20: 406-416.

Fox, G.E., E. Stackebrandt, R.B. Hespell, J. Gibson, J. Maniloff, T.A. Dyer, R.S. Wolfe, W.E. Balch, R.S. Tanner, L.J. Magrum, L.B. Zablen, R. Blackmore, R. Gupta, L. Bonen, B.J. Lewis, D.A. Stahl, K.R. Luehrsen, K.N. Chen and C.R. Woese 1980. The phylogeny of prokaryotes. *Science* 209: 457-463.

Garnery, L., D. Vautrin, J.M. Cornuet and M. Solignac 1991. Phylogenetic relationships in the genus *Apis* inferred from mitochondrial DNA sequence data. *Apidologie* 22: 87-92.

Gerbi, S.A. 1985. Evolution of ribosomal DNA, pp. 419-517. *In*

MacIntyre, R.J. ed., *Molecular Evolutionary Genetics*. Plenum, New York .

Gibson, G.A.P. and H. Goulet 1988. New insights for the phylogenetic reconstruction of Symphyta and the origin of Apocrita (Hymenoptera). *Proceedings of the XVIII International Congress of Entomology. Abstracts and Author Index*: 9.

Greenberg, B.D., J.E. Newbold and A. Sugino 1983. Intraspecific nucleotide sequence variability surrounding the origin of replication in human mitochondrial DNA. *Gene* 21: 33-49.

Gyllensten, U.B. and H.A. Erlich 1988. Generation of single-stranded DNA by the polymerase chain reaction and its application to direct sequencing of the HLA-DQA locus. *Proceedings of the National Academy of Sciences, USA* 85: 7652-7656.

Hamby, R.K. and E.A. Zimmer 1988. Ribosomal RNA sequences for inferring phylogeny within the grass family (Poaceae). *Plant Systematics and Evolution* 160: 29-37.

Hedges, S.B., K.D. Moberg and L.R. Maxson 1990. Tetrapod phylogeny inferred from 18S and 28S ribosomal RNA sequences and a review of the evidence for amniote relationships. *Molecular Biology and Evolution* 7: 607-633.

Hein, J. 1989. A new method that simultaneously aligns and reconstructs ancestral sequences for any number of homologous sequences, when the phylogeny is given. *Molecular Biology and Evolution* 6: 649-668.

Hein, J. 1990a. *Treealign Computer Program*. University of Montreal.

Hein, J. 1990b. Unified approach to alignment and phylogenies, pp. 626-644. *In* Doolittle R.F., ed. *Molecular Evolution: Computer Analysis of Protein and Nucleic Acid Sequences*. Methods in Enzymology. Vol, 183. Academic Press, New York.

Higgins, D.G. and P.M. Sharp 1988. CLUSTAL: a package for performing multiple sequence alignment on a microcomputer. *Gene* 73: 237-253.

Higgins, D.G., A.J. Bleasby, and R. Fuchs. 1992. CLUSTAL V: improved software for multiple sequence alignment. *Computer Applications in the Biosciences*.

Hillis, D.M. and S.K. Davis 1987. Evolution of the 28S ribosomal RNA gene in anurans: regions of variability and their phylogenetic implications. *Molecular Biology and Evolution* 4: 117-125.

Hillis, D.M. and M.T. Dixon 1989. Vertebrate phylogeny: evidence from 28S ribosomal DNA sequences, pp. 355-367. *In* Fernholm, B., K. Bremer and H. Jörnvall, eds. *The Hierarchy of Life* Elsevier, Amsterdam.

Hillis, D.M., A. Larson, S.K. Davis and E.A. Zimmer 1990. Nuceic Acids III: Sequencing, pp. 318-370. *In* Hillis, D.M. and C. Moritz, eds., *Molecular Systematics*. Sinauer Associates, Sunderland.

Hillis, D.M. and C. Moritz 1990. *Molecular Systematics*. Sinauer Associates, Sunderland.

Hixson, J.E. and W.M. Brown 1986. A comparison of the small ribosomal RNA genes from the mitochondrial DNA of the great apes and man: sequence, structure, evolution and phylogenetic implications. *Molecular Biology and Evolution* 3: 1-18.

HsuChen, C.-C. and D.T. Dubin 1984. A cluster of four transfer RNA genes in mosquito mitochondrial DNA. *Biochemistry International*. 8: 385-391.

HsuChen, C.-C., R.M. Kotin and D.T. Dubin 1984. Sequences of the coding and flanking regions of the large ribosomal subunit RNA gene of mosquito mitochondria. *Nucleic Acids Research* 12: 7771-7785.

Innis, M.A., D.H. Gelfand, J.J. Sninsky and T.J. White, eds. 1990. *PCR Protocols. A Guide to Methods and Applications*. Academic Press, New York.

Johnson, A.M. and P.R. Baverstock 1989. Rapid ribosomal RNA sequencing and the phylogenetic analysis of protists. *Parasitology Today* 5: 102-105.

Joss, J.M.P., N. Cramp, P.R. Baverstock and A.M. Johnson 1991. A phylogenetic comparison of 18S ribosomal RNA sequences of lungfish with those of other chordates. *Australian Journal of Zoology* 39: 509-518.

Jupe, E.R., R.L. Chapman, and E.A. Zimmer 1988. Nuclear ribosomal RNA genes and algal phylogeny - the *Chlamydomonas* example. *Biosystems* 21: 223-230.

Kocher, T.D., W.K. Thomas, A. Meyer, S.V. Edwards, S. Pääbo, F.X. Villablanca and A.C. Wilson 1989. Dynamics of mitochondrial DNA evolution in animals: amplification and sequencing with conserved primers. *Proceedings of the National Academy of Sciences, USA* 86: 6196-6200.

Königsmann, E. 1977. Das phylogenetische System der Hymenoptera. Teil 2: Symphyta. *Deutsche Entomologische Zeitung* 24: 1-40.

Ladiges, P.Y. and L.W. Martinelli, eds. 1990. *Plant Systematics in the Age of Molecular Biology*. CSIRO, Melbourne. (Also printed as Vol. 3 of *Australian Systematic Botany*.)

Lake, J.A. 1988. Origin of the eukaryotic nucleus determined by rate-invariant analysis of rRNA sequences. *Nature* 331: 184-186.

Lake, J.A. 1991. The order of sequence alignment can bias the selection of tree topology. *Molecular Biology and Evolution* 8: 378-385.

Lane, D.J., B. Pace, G.J. Olsen, D.A. Stahl, M.L. Sogin and N.R. Pace 1985. Rapid determination of 16S ribosomal RNA sequences for phylogenetic analyses. *Proceedings of the National Academy of Sciences, USA* 82: 6955-6959.

Larson, A. 1991. A molecular perspective on the evolutionary relationships of the salamander families. *Evolutionary Biology* 25: 211-277.

Larson, A. and A.C. Wilson 1989. Patterns of ribosomal RNA evolution in salamanders. *Molecular Biology and Evolution* 6: 131-154.

Lassner, M., O. Anderson and J. Bvorak 1987. Hypervariation associated with a 12-nucleotide direct repeat and inferences on intergenomic homogenization of ribosomal RNA gene spacers based on the DNA of a clone from wheat Nor-D3 locus. *Genome* 29: 770-781.

Li, W.-H., I. Wu and C.-C. Luo 1984. Nonrandomness of point mutations as reflected in nucleotide substitutions in pseudogenes and its evolutionary implications. *Journal of Molecular Evolution* 21: 58-71.

Maddison, D.R. 1991. The discovery and importance of multiple islands of most-parsimonious trees. *Systematic Zoology* 40: 315-328.

Maddison, W.P., M.J. Donoghue and D.R. Maddison 1984. Outgroup analysis and parsimony. *Systematic Zoology* 33: 83-103.

Maniatis, T., E.F. Fritsch and J. Sambrook 1982. *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor, New York.

Michener, C.D. 1990. Classification of the Apidae. *Kansas University Science Bulletin* 54: 75-164.

Mindell, D.P. 1991. Aligning DNA sequences: homology and

phylogenetic weighting, pp. 73-89. *In* Miyamoto M.M. and J. Cracraft, eds., *Phylogenetic Analysis of DNA Sequences*. Miyamoto M.M. and J. Cracraft, eds., Oxford University Press, New York.

Mindell, D.P. and R.L. Honeycutt 1989. Variability in the transcribed regions of ribosomal DNA and early divergences in birds. *Auk* 106: 539-548.

Mindell, D.P. and R.L. Honeycutt 1990. Ribosomal RNA in vertebrates: evolution and phylogenetic applications. *Annual Review of Ecology and Systematics* 21: 541-566.

Miyamoto, M.M. and S.M. Boyle 1989. The potential importance of mitochondrial DNA sequence data to eutherian mammal phylogeny, pp. 437-450. *In* Fernholm, B. K. Bremer and H. Jörnwall, eds. *The Hierarchy of Life*. Elsevier, Amsterdam.

Miyamoto, M.M. and J. Cracraft (eds.). 1991. *Phylogenetic Analysis of DNA Sequences*. Oxford University Press, New York.

Miyamoto, M.M., S.M. Tanhauser and P.J. Lailis 1989. Systematic relationships in the artiodactyl tribe Bovini (family Bovidae), as determined from mitochondrial DNA sequences. *Systematic Zoology* 38: 342-349.

Moritz, C., T.E. Dowling and W.M. Brown 1987. Evolution of animal mitochondrial DNA: relevance for population biology and systematics. *Annual Review of Ecology and Systematics* 18: 269-292.

Mullis, K.B. 1990. The unusual origin of polymerase chain reaction. *Scientific American* 262 (No. 4): 36-43.

Mullis, K., F. Faloona, S. Scharf, R. Saiki, G. Horn and H. Erlich 1986. Specific enzymatic amplification of DNA *in vitro*: the polymerase chain reaction. *Cold Spring Harbor Symposium on Quantitative Biology* 51: 263-273.

Mullis, K.B. and F.A. Faloona. 1987. Specific synthesis of DNA *in vitro* via a polymerase-catalyzed chain reaction. *Methods in Enzymology* 155: 335-351.

Needleman, S.B. and C.D. Wunsch 1970. A general method applicable to the search for similarities in the amino acid sequence of two proteins. *Journal of Molecular Biology* 48: 443-453.

Nei, M. 1987. *Molecular Evolutionary Genetics*. Columbia University Press, New York.

Pääbo, S. 1989. Ancient DNA: extraction, characterization, molecular cloning, and enzymatic amplification. *Proceedings of the National Academy of Sciences, USA* 86: 1939-1943.

Pääbo, S. 1990. Amplifying ancient DNA, pp.159-166. *In* Innis, MA., D.H. Gelfand, J.J. Sninsky and T.J. White, eds. *PCR Protocols*. Academic Press, New York.

Patterson, C. 1987. Introduction, pp. 1-22. *In* Patterson, C. ed. *Molecules and Morphology in Evolution: Conflict or Comprimise*. Cambridge University Press, Cambridge.

Powell, J.R., A. Caccone, G.D. Amato and C. Yoon 1986. Rates of nucleotide substitution in *Drosophila* mitochondrial DNA and nuclear DNA are similar. *Proceedings of the National Academy of Sciences, USA* 83: 9090-9093.

Qu, L.H., B. Michot and J.P. Bachellerie 1983. Improved methods for structure probing in large RNAs: a rapid 'heterologous' sequencing approach is coupled to the direct mapping of nuclease accessible sites. *Nucleic Acids Research* 11: 5903-5920.

Qu, L.H., N. Hardman, L. Gill, L. Chappell, M. Nicoloso and J.P. Bachellerie 1986. Phylogeny of helminths determined by rRNA sequence comparison. *Molecular and Biochemical Parasitology* 20: 93-99.

Rand, D.M. and R.G. Harrison 1989. Molecular population genetics of mitochondrial DNA size variation in crickets. *Genetics* 121: 551-569.

Rasnitsyn, A.P. 1980. [The origin and evolution of Hymenoptera]. *Trudy Paleontological Institute (Akademia Nauk) SSSR* 123: 1-191 (In Russian; English translation, 1984, Biosystematics Research Institute, Ottawa.)

Rasnitsyn, A.P. 1988. An outline of evolution of the hymenopterous insects (Order Vespida). *Oriental Insects* 22: 115-145.

Ross, H.H. 1937. A generic classification of the Nearctic sawflies (Hymenoptera, Symphyta). *Illinois Biological Monographs* 15: 7-173.

Saiki, R.K., S. Scharf, F. Faloona, K.B. Mullis, G.T. Horn, H.A. Erlich and N. Arnheim 1985. Enzymatic amplification of B-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. *Science* 230: 1350-1354.

Saiki, R.K., D.H. Gelfand, S. Stoffel, S.J. Scharf, R. Higuchi, G.T. Horn, K.B. Mullis and H.A. Erlich 1988. Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science* 239: 487-491.

Sakagami, S.F. and C.D. Michener 1987. Tribes of Xylocopinae and origin of the Apidae (Hymenoptera: Apoidea). *Annals of the Entomological Society of America* 80: 439-450.

Sanger, F., S. Nicklen and A.R. Coulson 1977. DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences, USA* 74: 5463-5467.

Sankoff, D. 1972. Matching sequences under deletion/insertion constraints. *Proceedings of the National Academy of Sciences, USA* 69: 4-6.

Sankoff, D., C. Morel and D.J. Cedergren 1973. Evolution of 5s RNA and the nonrandomness of base replacements. *Nature* 245: 232-234.

Schmickel, R.D., J. Sylvester, D. Stambolian and I.L. Gozalez 1990. The ribosomal DNA sequence for the study of human, chimpanzee and gorilla, pp. 11-32. *In* Dutta, S.K. and W.P. Winter, eds., *DNA Systmatics*, Vol. III. CRC Press.

Sellers, P. 1974. An algorithm for the distance between two finite sequences. *Journal of Combination Theory* 16: 253-258.

Sheppard, W.S. and B.A. McPheron 1991. Ribosomal DNA diversity in Apidae, pp. 89-102. *In* Smith, D.R. ed. *Diversity in the Genus Apis*. Westview Press, Boulder.

Smith, T.F., M.S. Waterman and W.M. Fitch 1981. Comparative biosequence metrics. *Journal of Molecular Evolution* 18: 38-46.

Smith, T.F., M.S. Waterman and C. Burks 1985. The statistical distribution of nucleic acid similarities. *Nucleic Acids Research* 13: 645-656.

Simon, C. 1991. Molecular systematics at the species boundary: exploring conserved and variable regions of the mitochondrial genome via direct sequencing from amplified DNA, pp. 33-72. *In* Hewitt, G.M., A.W.B. Johnston and J.P.W. Young, eds., *Molecular Techniques in Taxonomy*. NATO Advanced Series Institute, Springer Verlag, Berlin.

Simon, C., S. Pääbo, T.D. Kocher and A.C. Wilson 1990. Evolution of mitochondrial ribosomal RNA in insects as shown by the polymerase chain reaction, pp. 235-244. *In* Clegg, M.T. and S.J. O'Brien, eds., *UCLA Symposium on Molecular and Cellular Biology* New Series, Volume 122. Alan R. Liss, Inc., New York.

Simon, C., A. Franke and A. Martin 1991. The polymerase

chain reaction: DNA extraction and amplification, pp. 329-356. *In* Hewitt, G.M., A.W.B. Johnston and J.P.W. Young, eds., *Molecular Techniques in Taxonomy*. NATO Advanced Series Institute, Springer Verlag, Berlin.

Swofford, D.L. 1990. *PAUP: Phylogenetic Analysis Using Parsimony, Version 3.0L*. Illinois Natural History Survey, Champaign, Illinois.

Swofford, D.L. and G.L. Olsen 1990. Phylogenetic reconstruction, pp. 411-501. *In* Hillis D.M. and C. Moritz, eds., *Molecular Systematics*. Sinauer Associates, Sunderland.

Takahata, N. 1989. Gene geneology in three related populations: consistency probability between gene and populations trees. *Genetics* 122: 957-966.

Tautz, D., C. Tautz, D. Webb and G.A. Dover 1987. Evolutionary divergence of promoters and spacers in the rDNA family of four *Drosophila* species: implications for molecular coevolution in multigene families. *Journal of Molecular Biology* 195: 525-542.

Thomas W.K. and A.T. Beckenbach 1989. Variation in salmonid mitochondrial DNA: evolutionary constraints and mechanisms of substitutions. *Journal of Molecular Evolution* 29: 233-245.

Vawter, L. 1991. Evolution of blattoid insects and of the small subunit ribosomal RNA gene. Doctoral Dissertation, University of Michigan.

Vawter, L. and W.M. Brown 1986. Nuclear and mitochondrial DNA comparisons reveal extreme rate variation in the molecular clock. *Science* 234: 194-196.

Vlasak, I., S. Burgschwaiger and G. Dreil 1987. Nucleotide sequence of the large ribosomal RNA of honey bee mitochondria. *Nucleic Acids Research* 15: 2388.

Vossbrinck, C.R. and S. Friedman 1989. A 28S ribosomal RNA phylogeny of certain cyclorrhaphous Diptera based upon a hypervariable region. *Systematic Entomology* 14: 417-431.

Walker, W.F. 1985. 5S and 5.8S ribosomal RNA sequences and protist phylogenies. *Biosystems* 18: 269-278.

Waterman, M.S., T.F. Smith and W.A. Bayer 1976. Some biological sequences metrics. *Advances in Mathematics* 20: 367-387.

Waterman, M.S., J. Joyce and M. Eggert 1991. Computer alignment of sequences. pp. 59-72. *In* Miyamoto M.M. and J. Cracraft, eds. *Phylogenetic Analysis of DNA Sequences*. Oxford University Press, New York.

Watrous, L.E. and Q.D. Wheeler 1981. The outgroup comparison method of character analysis. *Systematic Zoology* 30: 1-11.

Wheeler, W.C. 1989. The systematics of insect ribosomal DNA, pp. 307-321. *In* Fernholm, B., K. Bremer and H. Jörnvall, eds. *The Hierarchy of Life* Elsevier, Amsterdam. Amsterdam.

Wheeler, W.C. 1990. Nucleic acid sequence phylogeny and random outgroups. *Cladistics* 6: 363-367.

Wheeler, W.C. and R.L. Honeycut 1988. Paired sequence difference in ribosomal RNAs: evolutionary and phylogenetic implication. *Molecular Biology and Evolution* 5: 90-96.

Woese, C.R. 1987. Bacterial Evolution. *Microbiological Reviews* 51: 221-271.

Woese, C.R. , O. Kandler and M.L. Wheelis 1990. Towards a natural system of organisms: proposal for the Domains Archaea, Bacteria and Eucarya. *Proceedings of the National Academy of Sciences, USA* 87: 4576-4579.

Zimmer, E.A., R.K. Hamby, M.L. Arnold, D.A. Leblanc and E.L. Theriot 1989. Ribosomal RNA phylogenies and flowering plant evolution, pp. 205-214. *In* Fernholm, B., K. Bremer and H. Jörnvall, eds. *The Hierarchy of Life*. Elsevier, Amsterdam.

APPENDIX 1. MITOCHONDRIAL DNA PRIMERS

The following primers are based on hymenopteran mtDNA sequences and have been successfully employed on a range of species (Cameron 1991; Cameron unpublished data). All primers are written in the 5' to 3' direction. Primers are named for the gene in which they are located (e.g. 16S), their relative position in the gene (m=mid, l=low), and whether they prime in the forward (F) or reverse (R) direction. Two nucleotides at a single position (one below the other) represent a degenerate site (a nucleotide site occupied by more than one nucleotide). Degeneracy in the primer allows for some degree of mismatch between the primer and its complementary target.

16S rRNA Primers	
875-16SmF (24mer)	
Apis	5'-TTATTCACCTGTTTATCAAAACAT-3'
874-16SlR (20mer)	
Apis	5'-TATAGATAGAAACCAATCTG-3' C
16SmR (20mer)	
Cotesia (Braconidae)	5'-CAGGTGAATATAAATTTGCC-3'
12S rRNA Primers	
12SmF (20mer)	
Bombus	5'-CTTATTAGAGAAACTTGTAG-3'